DOCUMENT RESUME

ED 324 353 TM 015 610

AUTHOR Cohen, Michael P.
TITLE Error Analysis of Estimates for Less-Than-Two-Year Postsecondary Education Institutions.
PUB DATE 88
NOTE 6p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Data Analysis; Databases; *Data Collection; *Estimation (Mathematics); *Institutional Characteristics; *National Surveys; Nonprofit Organizations; Postsecondary Education; Private Schools; Public Schools; Research Methodology; Research Problems; *Sampling; Statistical Analysis
IDENTIFIERS *Error Analysis (Statistics); Integrated Postsecondary Education Data System; *Less Then Two Year Postsecondary Institutions

ABSTRACT

The Integrated Postsecondary Education Data System (IPEDS) of the National Center for Education Statistics collects data from all sections of postsecondary education. For less-than-two-year institutions, the IPEDS collects data on fall enrollment, completions (earned awards and certificates), etc., from all public institutions in that category and from a sample of private institutions. The methodological issues that arise from trying to collect data from less-than-two-year postsecondary education institutions are discussed. The variable nature of the universe of these institutions, which go in and out of business frequently, contributes to errors in IPEDS estimates. The IPEDS sectors discussed are: (1) less-than-two-year public institutions; (2) less-than-two-year private non-profit institutions; and (3) less-than-two-year private for-profit institutions. Sampling errors were computed by balanced repeated replication based on 32 replications. Data are provided for 330,007 male and 425,835 female students attending the sample institutions in the fall of 1987 and for 311,561 male and 378,267 female students who graduated from the sample institutions in 1987-88. Challenging problems are presented to IPEDS statisticians because of: reluctance of the institutions to respond to the survey; the volatile nature of the population; the highly skewed distribution of the sizes of the institutions; and the sensitivity of the data to precise definitions. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED324353

ERROR ANALYSIS OF ESTIMATES FOR LESS-THAN-TWO-YEAR POSTSECONDARY EDUCATION INSTITUTIONS

Michael P. Cohen
National Center for Education Statistics

TM015610

ERROR ANALYSIS OF ESTIMATES FOR LESS-THAN-TWO-YEAR POSTSECONDARY EDUCATION INSTITUTIONS

Michael P. Cohen, National Center for Education Statistics*
555 New Jersey Avenue NW, Washington, DC 20208-5654

Key Words: Sampling, nonsampling, response rate, universe

Abstract: The Integrated Postsecondary Education Data System (IPEDS) of the National Center for Education Statistics collects data from all sectors of postsecondary education. For less-than-two-year institutions, data are sought on fall enrollment, completions, etc., from all public institutions in the universe and from a sample of private institutions (both nonprofit and for-profit institutions). In addition to sampling error, additional sources of error arise because of the highly variable nature of the universe of less-than-two-year institutions that move and go in and out of business frequently. Issues also arise in precisely defining postsecondary education. This paper explores these issues and analyses their effects upon the IPEDS estimates.

Introduction:

Less-than-two-year postsecondary education institutions (public, private nonprofit, and private for profit) provide a significant portion of the postsecondary education in the nation. Furthermore, students in these institutions receive a substantial share of federal student financial aid. The less-than-two-year postsecondary education sectors constitute the biggest increase in coverage of the Integrated Postsecondary Data System (IPEDS) of the National Center for Education Statistics (NCES) over the previous NCES higher education institution-level surveys. This paper explores the methodological issues that arise in trying to collect data from these important sectors.

Background on the IPEDS:

The NCES has established the Integrated Postsecondary Education Data System (IPEDS) as its core postsecondary education data collection program. It is a single, comprehensive system that encompasses all identified institutions whose primary purpose is to provide postsecondary education. IPEDS supersedes the Higher Education General Information Survey (HEGIS). HEGIS collected data from 1965-1986 from the universe of institutions accredited as an institution of higher education by an organization recognized by the Secretary of Education.

In designing and implementing IPEDS, a significant effort has been made to maintain continuity with the HEGIS data series. That is, IPEDS is based on the HEGIS model in that institution level data are collected either directly from the institution or through a central, State coordinating office. All HEGIS institutions are included in IPEDS. However, IPEDS also includes non-accredited institutions of postsecondary education as well as proprietary institutions -- institutions not found in the HEGIS database. As a result, IPEDS includes approximately 12,000 postsecondary institutions whereas only approximately 3,500 accredited colleges and universities were included in HEGIS. (Accreditation is not a requirement for inclusion in IPEDS as was the case with HEGIS: rather, accreditation is a characteristic about which data are collected in IPEDS.)

IPEDS includes the following institutions of postsecondary education: baccalaureate or higher institutions, two-year award institutions, and less-than-two-year institutions (that is, institutions whose awards usually result in terminal occupational awards or are creditable toward a formal two-year or higher award). Each of these three categories is further disaggregated by control (public, private nonprofit, private for profit) resulting in nine institutional categories or sectors.

IPEDS employs several new concepts in the collection of postsecondary institutional educational data. Data elements are both relevant to all providers of postsecondary education and consistent among components of the system that have been formulated and tested. A set of data elements has been established to identify characteristics common to all providers of postsecondary education. Specific data elements have been established to define unique characteristics of different types of providers of postsecondary education. Relationships and dependencies among several components of IPEDS have been established to avoid duplicate reporting and enhance the policy relevance and analytic potential of the data. IPEDS has also recognized the problems involved in trying to make inter-State and inter-institutional comparisons using the NCES postsecondary data and has addressed many of these problems through the use of clarifying questions -- questions that ask what was or was not included in a reported count or total. Finally, specialized but compatible reporting formats have been developed for the different sectors of postsecondary education providers. In general, the reports developed for postsecondary institutions granting baccalaureate and higher degrees are the most extensive: forms for the two-year, and less-than-two-year award granting sectors request less data. This design feature accommodates the varied operating characteristics, program offerings, and reporting capabilities that differentiate postsecondary institutional sectors while, at the same time, yielding comparable statistics for items common to all postsecondary sectors.

Definition of Postsecondary Education:

NCES defines postsecondary education as the provision of formal instructional programs whose curriculum is designed primarily for students who have completed the requirements for a high school diploma or its equivalent. This is to include programs whose purpose is academic, vocational, and continuing education, and to exclude avocational and adult basic education programs.

Scope of IPEDS:

NCES recognizes that postsecondary education is conducted by a variety of providers; however, IPEDS is limited to those institutions (or subsidiary elements of an institution) whose primary purpose is the provision of postsecondary education. The following providers are included in the conceptual framework underlying the IPEDS data collection:

TM015610

1. Baccalaureate or Higher Institutions

   Institutions or subsidiary elements whose primary purpose is the provision of postsecondary education and that award at least a baccalaureate or higher degree in one or more programs.

2. Two-Year Award Programs

   Institutions or subsidiary elements whose primary purpose is the provision of postsecondary education and that confer at least a two-year formal award (certificate or associate degree) or have a two-year program that is creditable toward a baccalaureate or higher degree in one or more programs but do not award a baccalaureate degree.

3. Less-Than-Two-Year Institutions

   Institutions or subsidiary elements whose primary purpose is the provision of postsecondary education and whose programs are less than two years in duration and result in a terminal occupational award or are creditable toward a formal two-year or higher award.

Each of these categories (called levels) is further disaggregated by control (public, private nonprofit, and private for profit) resulting in nine institutional categories or sectors. This paper discusses only sector 7 (Less-Than-Two-Year Public Institutions), sector 8 (Less-Than-Two-Year Private Nonprofit Institutions), and sector 9 (Less-Than-Two-Year Private For Profit Institutions).

Components of IPEDS:

IPEDS consists of several integrated components that obtain information on who provides postsecondary education (institutions), who participates in it and completes it (students), what programs are offered and what programs are completed, and the resources involved in the provision of institutionally based postsecondary education, both human resources and financial resources. Specifically, those components that pertain to the less-than-two year sectors and their standard abbreviations are Institutional Characteristics (IC), Total Institutional Activity (EA), Fall Enrollment (EF), Fall Enrollment in Occupationally Specific Programs (EP), Completions (C), Finance (F), and Staff (S).

The Institutional Characteristics (IC) survey is the only IPEDS survey that is sent to all institutions in the IPEDS universe. The other surveys are sent to the universe of four-year and two-year institutions, to all public less-than-two-year institutions, and to a sample of private less-than-two-year institutions. For example, the 1987 Fall Enrollment (EF) survey was sent to 2,752 four-year institutions, 2,985 two-year institutions, 481 public less-than-two-year institutions, and to a sample of 1,223 (out of 6,234) private less-than-two-year institutions. The 6,234 private less-than-two-year institutions constitute over half the entire universe of 12,070 institutions. However, these institutions typically have relatively small student enrollments at any given time.

Sample of Private Less-Than-Two-Year Institutions:

In this section, the 1987 IPEDS private less-than-two-year sample will be described. As noted above, this sample determines which private less-than-two-year institutions will receive the various IPEDS survey forms with the exception of the Institutional Characteristics form, which is mailed to all institutions in the universe. The 1987 sampling techniques were a refinement of the methods used to select the 1986 IPEDS sample, the first such sample selected for IPEDS. The 1988, 1989, and 1990 samples have been selected based on the same methodology. The information in this section is based on the documentation of ORI, Inc. (1986) and Cohen (1987).

Throughout this section, the institutions discussed should be understood to be less-than-two year private institutions only. A particular subset of these institutions, referred to as "HEGIS" institutions, require some special discussion. HEGIS institutions are institutions that would meet the definition for inclusion in the old HEGIS surveys (recall that HEGIS was the predecessor to IPEDS); namely these institutions are accredited as institutions of higher education by an organization recognized by the Secretary of Education. In order to maintain continuity of IPEDS with HEGIS, HEGIS institutions are selected with certainty (that is, they are self-representing) regardless of which sector they are in.

As a first step in selecting the sample, each of the 22 private less-than-two-year HEGIS institutions was placed in the sample as a certainty selection. In addition, all 1986 sampled institutions, still in-scope, were retained. There were 846 such (non-HEGIS) institutions. Another 355 institutions, not previously sampled, were added to the 1987 sample. The sample was selected from sector 8 (less-than-two year private nonprofit) and sector 9 (less-than-two- year private for-profit). Each sector was divided into strata and, after sorting by zip code, systematic (interval) sampling was used within a stratum.

Sampling Frame:

The sampling frame used for this sample was the IPEDS universe current as of July 2, 1987, and contained 6,372 less-than-two-year private institutions. Of these, 599 were in sector 8 and 5,773 were in sector 9. Of the sector 8 institutions, 2 were in HEGIS and 70 others were in the 1986 sample. Of the sector 9 institutions, 20 were in HEGIS and 776 others were in the 1986 sample. (These counts include institutions that changed sector from 8 to 9 or 9 to 8. Institutions that had been classified in sectors 1 to 7 are not considered to be in the 1986 sample. Institutions that are no longer classified in sectors 8 or 9 but were in 1986 are not among the 6,372 institutions in the sector 8 and 9 universe.)

Stratification:

Each sector was divided into four strata. All HEGIS institutions (identified on the data set by having a FICE code) were assigned to stratum 3 (the numbering of the strata was chosen for consistency with the sampling procedures report for the 1986 sample). Stratum 1 consisted of those institutions which responded to the 1985-86 Institutional Characteristics (IC) survey, and stratum 2 consisted of those institutions in the 1985-86 survey which did not respond. Stratum 4, the new stratum, consisted of those institutions which were not in either sector 8 or 9 of the 1985-86 IC survey.

Originally, stratum 4 had been divided into two strata: stratum 4a for institutions that responded to the 1986-87 IC survey and stratum 4b for those institutions that did not

respond. However, there was too little enrollment information for stratum 4b so the two strata were combined.

Enrollment Data:

The square root of enrollment was chosen as the measure of size for sample selection. The enrollment figure from the IC survey was used for those institutions that responded to this item. For the other institutions, the enrollments were imputed. The procedure used to impute the enrollment data depended on whether or not the institution was in the fall enrollment (EF) survey and responded to total enrollment for EF.

1. If the EF enrollment was present but IC enrollment was missing, then the EF enrollment was adjusted and used to compute the imputed size measure. The procedure consisted of two steps:

(a) The ratios r

$$r = \frac{\text{Sum of } e_{ic}}{\text{Sum of } e_{ef}}$$

were computed for each stratum in each sector, where

e_{ic} = enrollment from IC

e_{ef} = enrollment from EF

and the sums in the ratio are over those institutions in the stratum and sector that reported both IC and EF enrollment data.

(b) The size measure for the institution is then

size = square root of $(r \cdot e_{ef})$

where e_{ef} is the enrollment from EF for the institution in question.

This adjustment procedure is used to reflect reporting differences in enrollment between the IC and EF surveys. In general, IC enrollment tends to be somewhat higher than EF enrollment. An estimate of the IC enrollment could be derived by squaring the imputed size.

2. If neither EF nor IC enrollment was present for the institution, its measure of size was imputed by the mean of the measures of size for all institutions in the stratum and sector that reported enrollment for either IC or EF (thus, the measures of size computed in process 1 were used in the computation).

Note that because the square root of enrollment is the measure of size, the square root of enrollment rather than the enrollment itself was used throughout the imputation calculations. Once these imputation procedures had been completed, every institution had associated with it a measure of size.

Stratum Sample Sizes:

For stratum 3 (HEGIS institutions), each school was selected with certainty. There were 2 in sector 8 and 20 in sector 9.

For the other strata (non-HEGIS institutions), the sample size for each stratum and sector was determined proportional to the measure of size (square roots of enrollment) of the institutions in that stratum and sector. Because of the decision to retain institutions from the 1986 sample, the number of new selections had to be reduced by the appropriate amount. The total sample size for the non-HEGIS institutions was set to be approximately 1,200 (in fact, it was 1,201). A summary of the sample sizes by sector and stratum is given in the following tables.

	Sector 8		
	retained	new selection	total
stratum 1	22	40	62
stratum 2	48	15	63
stratum 4	0	3	3
total for sector (non-HEGIS)	70	58	128

	Sector 9		
	retained	new selection	total
stratum 1	208	248	456
stratum 2	568	4	572
stratum 4	0	45	45
total for sector (non-HEGIS)	776	297	1073

Sampling Method:

The HEGIS institutions and retained institutions from 1986 were included in the sample first. They were then removed from the list of institutions for probability selection.

Sampling was independent by sector and stratum with the probability of selection proportional to the measure of size. Institutions were sorted by zip code within sector and stratum to insure geographic dispersion. Systematic (interval) sampling was used with a new random start for each stratum within each sector.

Three institutions had measures of size greater than the sampling interval. This means that these institutions had a positive probability of being selected more than once for the sample ("multiple hits"). Had an institution been selected more than once, a weighting adjustment would have been performed to account for the multiple hit. However, in fact, no multiple hits occurred.

Survey Response:

The following data on the response rates to the 1987 IPEDS Fall Enrollment Survey for less-than-two-year institutions, taken from Kroe (1988), illustrate the volatility of these sectors.

	Public	Private
Institutions receiving form	481	1,223
Institutions declared out-of scope	113	175
In-scope institutions receiving survey form	368	1,048
In-scope institutions responding to form	173	801
(Unweighted) response rate	47.0	76.4

For the private less-than-two-year institutions, more detail can be provided. Sectors 8 and 9 denote respectively the nonprofit and for-profit less-than-two-year institutions. The four sample strata are described in the sections on sampling.

Sector	sample stratum	sample size sample size	sample size after deletions	Institutions responding	Response rate
Total	-	1,223	1,041*	798	76.7
8	1	62	47	37	78.7
8	2	63	45	33	73.3
8	3	2**	0	0	-
8	4	3	3	2	66.7
9	1	456	420	341	81.2
9	2	572	487	358	73.5
9	3	20**	8	5	62.5
9	4	45	31	22	71.0

*After the sample was drawn, an additional 7 institutions in the Fall Enrollment universe were reclassified into the private less-than-two-year sectors. Three of these institutions were respondents, and two were HEGIS institutions.

**During the processing of the 1987 sample enrollment data, 14 of the 22 HEGIS institutions that were selected with certainty were no longer in-scope as private less-than-two-year institutions. In addition, 4 responding institutions selected as part of the sample were identified as HEGIS schools after the sample was drawn. Thus, 12 HEGIS institutions are in the sample.

Sampling Error:

The sampling errors for estimates have been computed by the method of balanced repeated replication based on thirty-two replicates. The nonresponse adjustment was performed at the replicate level in addition to the full sample level. To give some notion of the magnitude of the sampling errors, they are presented below for two specific cases: (1) 1987 Fall Enrollment estimates by sector for men, women, and total, and (2) 1987-88 Completions (earned awards and certificates) by level of award for men, women, and total.

Less-Than-Two-Year Postsecondary Institutions
1987 Fall Enrollment Estimates and Sampling Error

		estimate	standard error	coefficient of variation
Public	men	70,948	0	0
	women	55,433	0	0
	total	126,381	0	0
Private Nonprofit	men	22,025	6,194	.281
	women	19,605	5,679	.290
	total	41,630	11,327	.272
Private For-Profit	men	237,034	54,282	.229
	women	350,797	43,569	.124
	total	587,831	94,164	.160
All Less-Than-Two-Year	men	330,007	54,361	.165
	women	425,835	45,137	.106
	total	755,842	95,756	.127

Note that the sampling errors for estimates for public less-than-two-year institutions are zero because all such institutions in the universe are sent the Fall Enrollment forms. Data for nonrespondents are imputed, which is one source of nonsampling error for these institutions.

Less-Than-Two-Year Postsecondary Institutions
1987-88 Completions Estimates and Sampling Error

		estimate	standard error	coefficient of variation
Less-Than-One-Year Awards	men	270,528	43,282	.160
	women	321,614	37,604	.117
	total	592,142	73,449	.124
At Least One But Less-Than-Two-Year Awards	men	41,033	10,306	.251
	women	56,653	5,204	.092
	total	97,686	14,243	.146
All Less-Than-Two-Year Awards	men	311,561	43,742	.140
	women	378,267	38,654	.102
	total	689,827	75,037	.109

As one may suspect from the relatively large coefficients of variation, the distributions of fall enrollment and completions (and other statistics) from the private less-than-two-year institutions are extremely skewed. For example, the five highest reported 1987 fall enrollments by these institutions were 13,000, 12,000, 7,702, 2,876, and 2,332 whereas the estimated median fall enrollment per institution was only 55.

Discussion:

The need to obtain statistics from the less-than-two-year postsecondary education institutions presents many challenging problems to the survey statistician. The reluctance of the institutions to respond, the volatile nature of the population, the highly skewed distribution of sizes of the institutions, the sensitivity of the data to precise definitions, among other issues, all contribute to the difficulties. Nevertheless, these institutions are an interesting and important part of the education enterprise about which we need to know more.

References

Cohen, Michael P. (1987), "1987 IPEDS Sample -- Sectors 8 and 9," NCES, Unpublished.

Kroe, P. Elaine (1988), "Documentation for Data Tape: 1987 Fall Enrollment in Postsecondary Institutions," NCES Publication 88-226, U.S. Department of Education.

ORI, Inc. (1986), "Sampling Procedures Report for the 1986 IPEDS Less-Two-Year Sample," Memorandum to NCES.

*This paper is intended to promote the exchange of ideas among statisticians. The views are those of the author, and no official support by the U.S. Department of Education is intended or should be inferred.